

Geocoding EPHT Data

Thomas Talbot

Environmental Health Surveillance Section
NYS Department of Health

Geocoding turns data records containing address or location information into geographic objects that can be displayed on a map.

Types of data which is geocoded

- Residential addresses
- Schools, hospitals
- Emission sources, hazardous waste sites
- Wells, air monitoring stations

Examples

- **Street address:**
 - 118 Lincoln Ave, Altamont, NY 12009
- **Other location information:**
 - World Trade Center
 - Wood Lake Mobile Home Park
 - Yankee Stadium
 - ZIP or ZIP+4

Key Elements of Street Geocoding

- Table of street addresses to match
 - 118 Lincoln Ave. 12009
- Reference file to match with, includes range of street numbers, street name, ZIP & geographic coordinates.
- Software to relate the two tables and assign coordinates.
- People to check for errors and match rejects.

Geocoding Process

- Standardizes addresses
- Finds Street within area (ZIP Code)
- Finds street segment within the range
- Determines side of street by parity of street number

Geocoding Process #2

- Interpolates between nodes
- Adds an offset from the road (~30 ft)
- Record is linked to geographic object (point)

Geocoding Street Addresses

- Address is split into fields –
 - Street number, street name, town, ZIP

Number	Street	Type	Other	Town	ZIP
1209	ROUTE 20		SUITE A	GUILDERLAND	12456
118	LINCOLN	AVENUE	APT 4	ALTAMONT	12009
37	W MAIN	STREET		TROY	12567

Address Standardization

Purpose: Produce Single, Standard Address

123 Oak Street West



123 Oak ST W

Fifth Avenue 2105



2105 Fifth AV

83 La Gata Lane



83 La Gata LA

Standard Street Type Abbreviations

CR	CR	FWY	FRWY
Crescent	CR	Highway	HWY
Crescent	CR	HWY	HWY
CRSG	CRSG	HY	HWY
CSWY	CSWY	LA	LA
DR	DR	Lane	LA
Drive	DR	LN	LA
DRV	DR	Loop	Loop
EXP	EXWY	Motorway	MTWY
Expressway	EXWY	MTWY	MTWY
EXPWY	EXWY	Parkway	PKWY
Freeway	FRWY	PKWY	PKWY
FRWY	FRWY	etc.	

Alias Tables

Alias	Address
CMU	5000 Forbes Av
Carnegie Mellon	5000 Forbes AV
Carnegie Mellon Univ	5000 Forbes AV
Carnegie Mellon University	5000 Forbes AV

**Once addresses are standardized
they can be matched to
standardized reference files which
have geographic coordinates.**

Assigning a Coordinate

Address

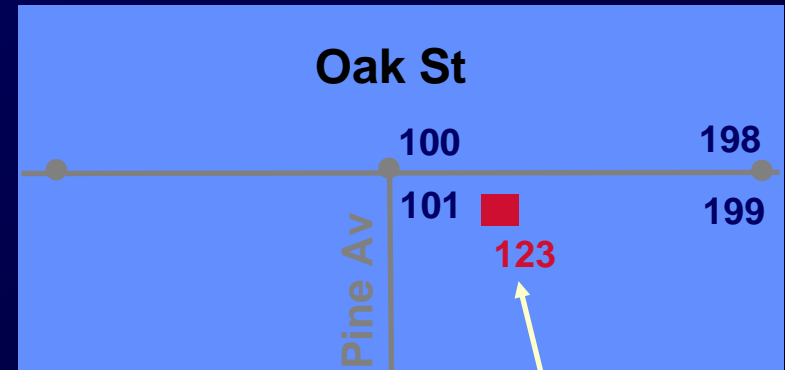
123 Oak St

333 Pine Av

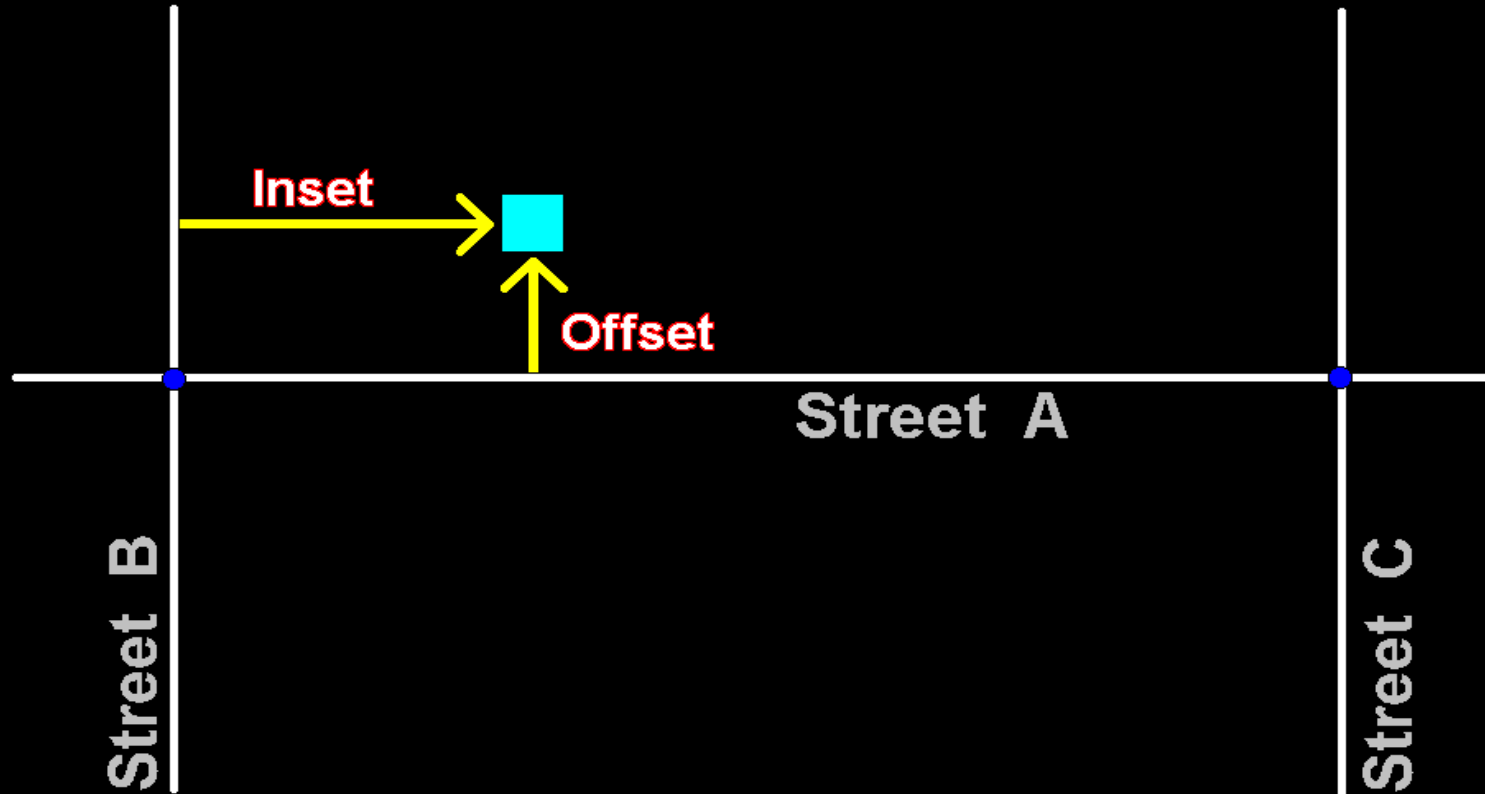
18 Bay Dr

Streets

From	To	Street	Type	Side	Parity
2	98	Oak	St	L	E
1	99	Oak	St	R	O
100	198	Oak	St	L	E
101	199	Oak	St	R	O



Offset & Inset Adjustment



Geocoding Result Codes or Score

- Represents success or failure of geocoding
- Information on the quality of the match.
- Code.
 - S5HPNTSZ
- Score
 - 80

Problems:

- Addresses may be duplicated
Two 118 Lincoln Aves in Albany County
- Variations in Street Names
Western Ave, Route 20
- Data Entry Errors
Fidth Avenue
Maine vs. Main
- Intersections
5th AV and Craig St

Problems

- Rural Routes, Post Office Boxes
- Mobile home parks
- Apartment Complexes
- Street Ranges incorrect in reference file

More Problems

- **Renumbering of streets**

E911

- **Street Number problems**

147 ½ Pine Court

- **Abbreviations**

BNX → Bronx

- **Extra Info**

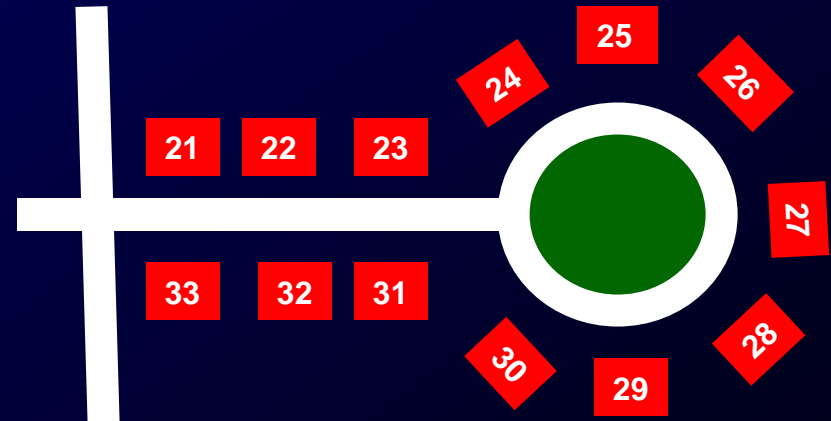
38 Maple Ave Apt 5B

Problems with Parity

Digital Reference File

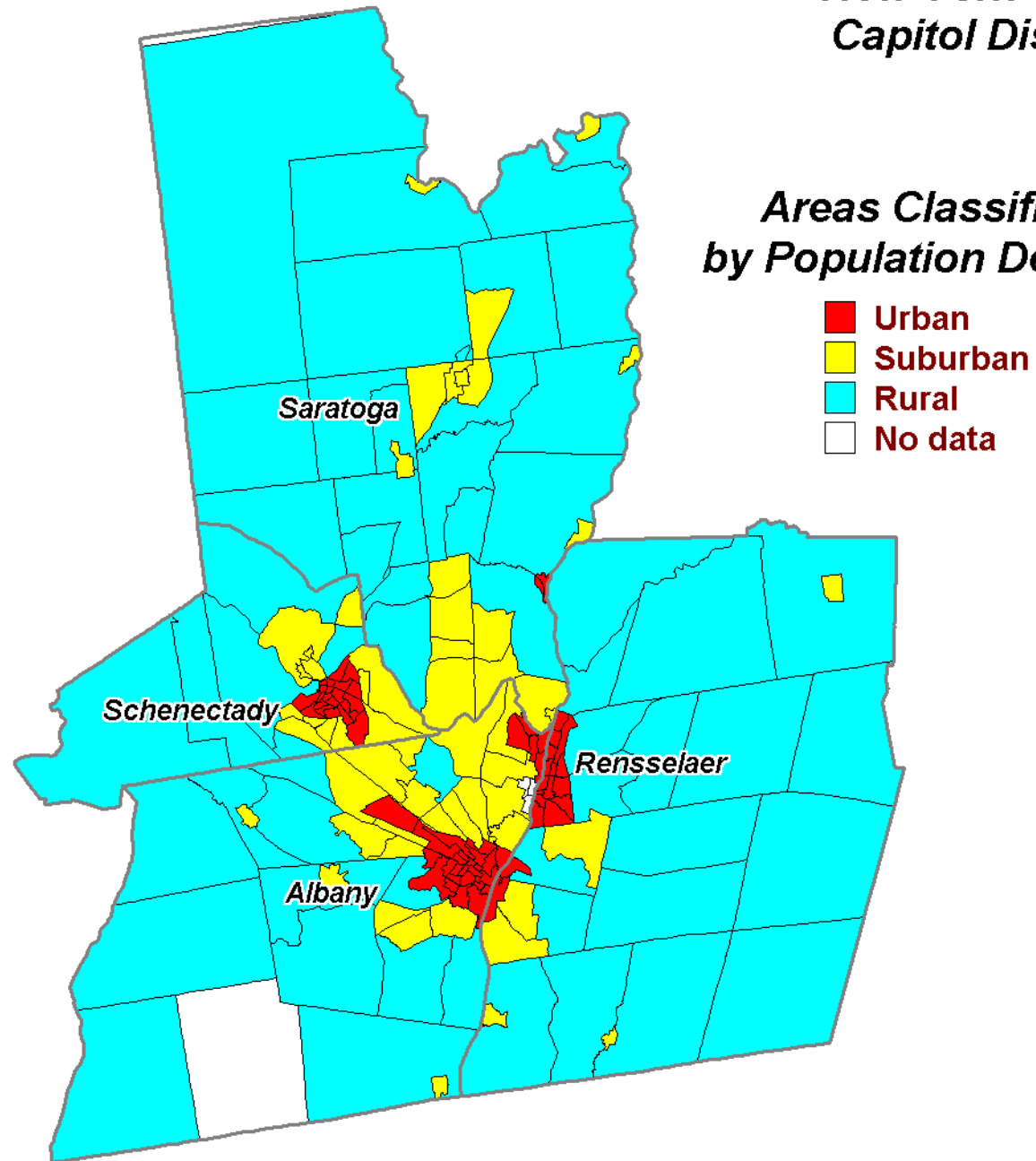
21	33
20	32

Actual



New York State Capitol District

Areas Classified by Population Density



Automated Geocoding Match Rates NY Capitol Region

Urban	94%
--------------	------------

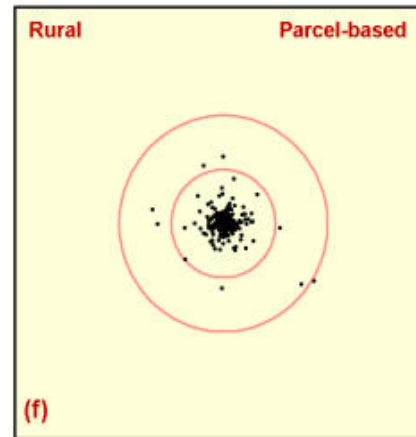
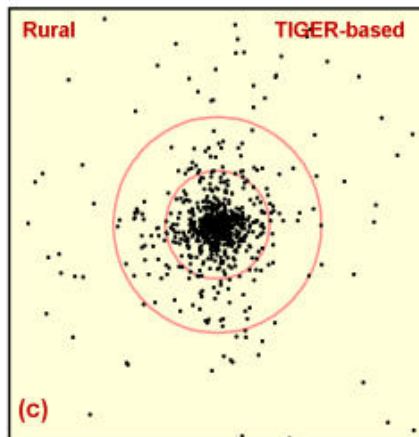
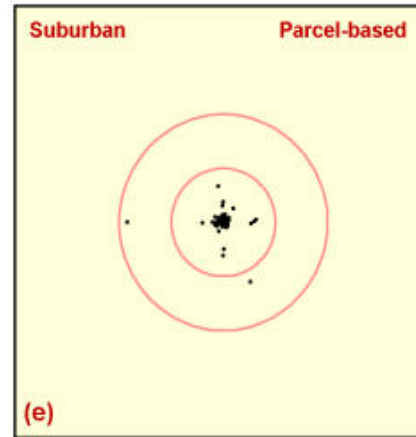
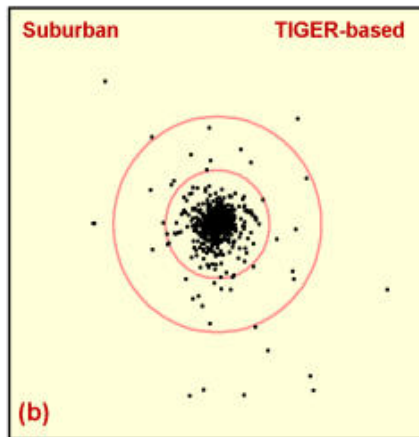
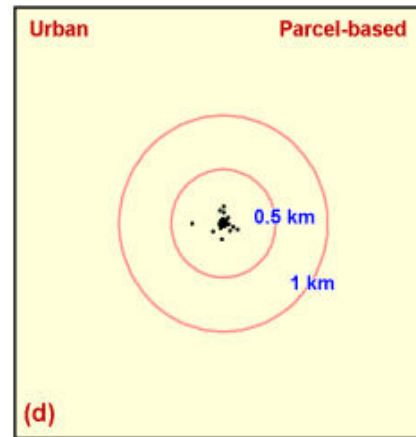
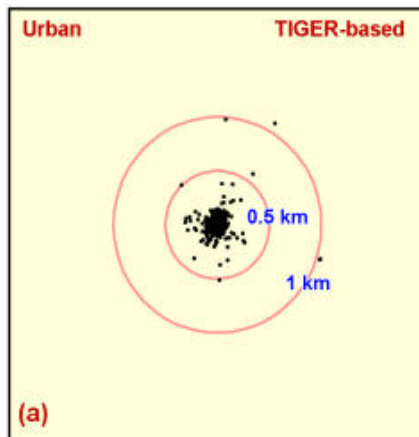
Suburban	87%
-----------------	------------

Rural	62%
--------------	------------

Positional error automated geocoding

New York State Capitol Region

Area	Accuracy (meters)	
	75%	95%
Urban	62	152
Suburban	158	421
Rural	498	2,872

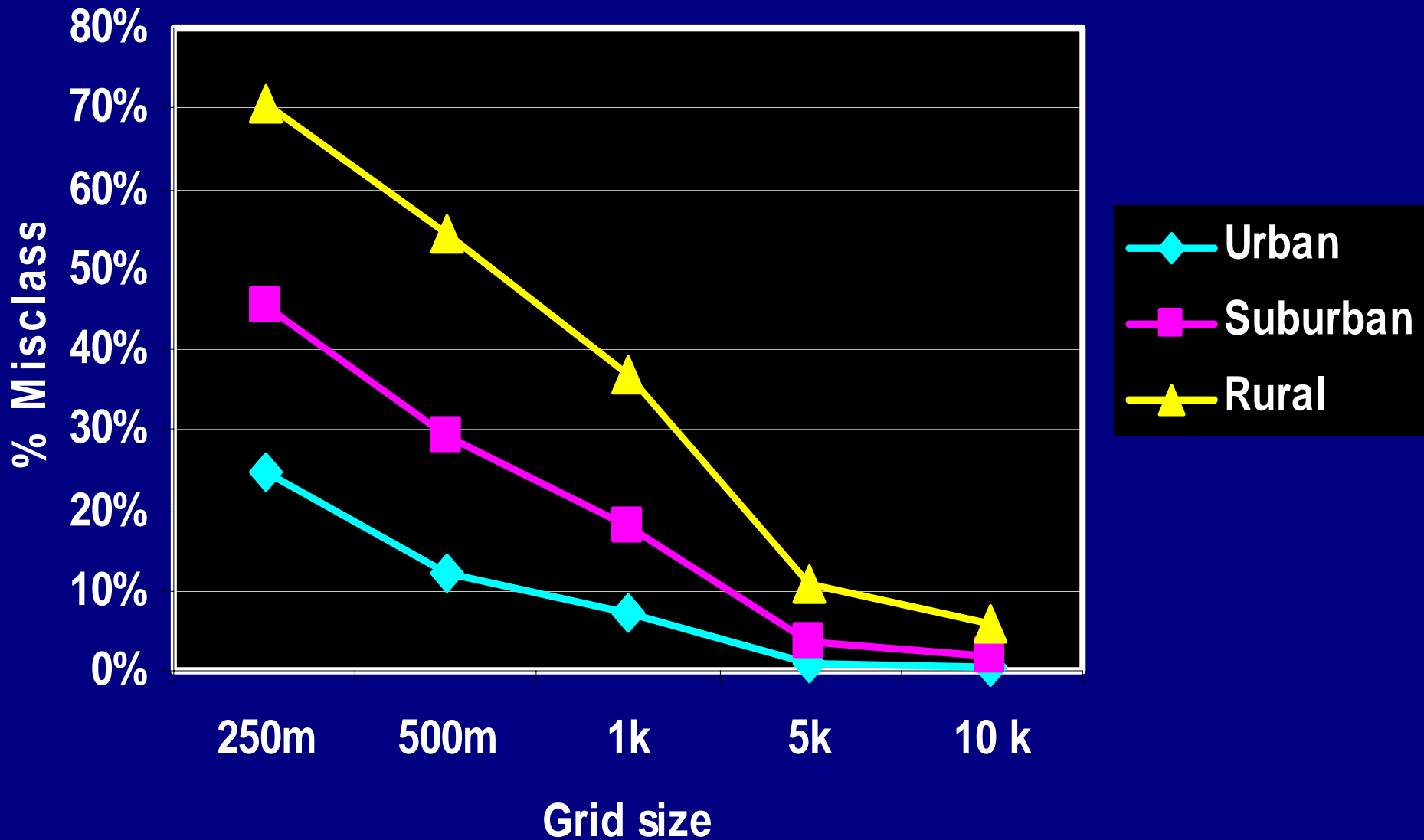


Geocoding Accuracy in New York Real Property vs. TIGER Based

	95% Accuracy (meters)	
	Parcels Centroids	TIGER
Urban	21	152
Suburban	39	421
Rural	195	2,872

Cayo & Talbot, 2003

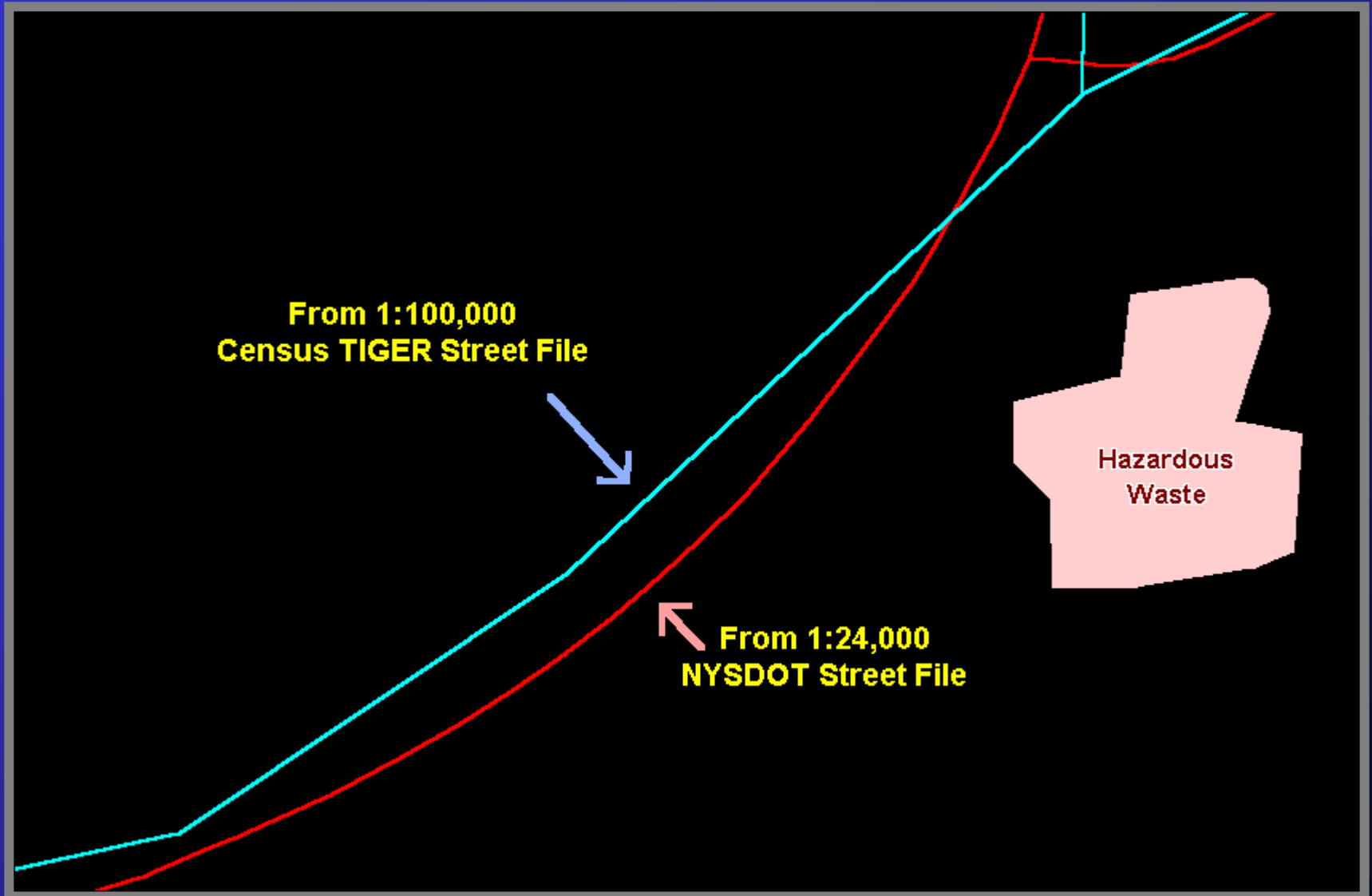
Misclassification of geocoded points in a Grid



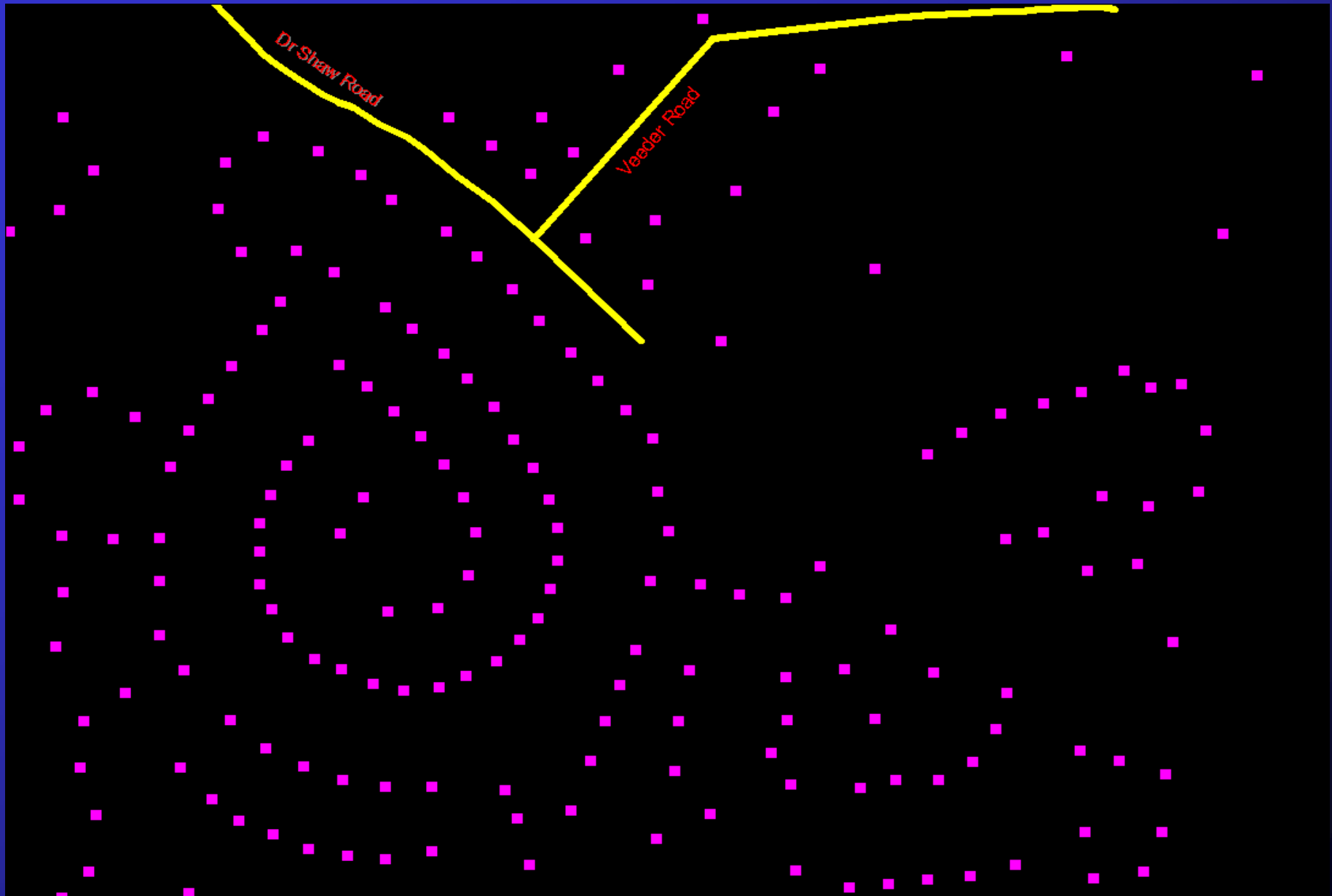
Disparities between the Census TIGER files and Local Street Files

- **More accurate reference street files can cause cases to be geocoded in the correct location but in the wrong census area.**
- **More accurate street files need to be incorporated into the census database.**

Which is the best file for Geocoding



New Addresses not in Street File



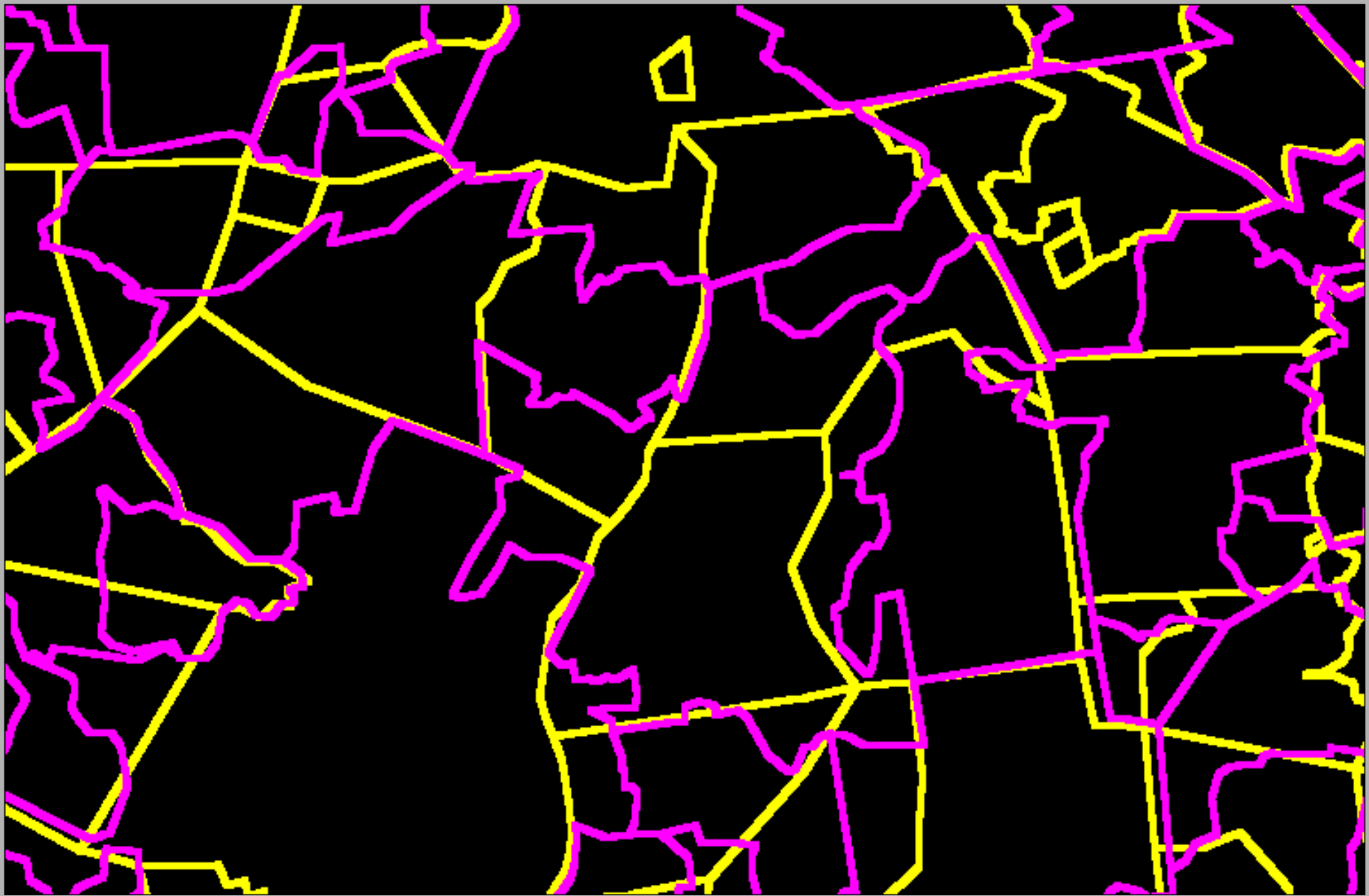
1990 GDT ZIP Code Boundaries



1991 GDT ZIP Code Boundaries

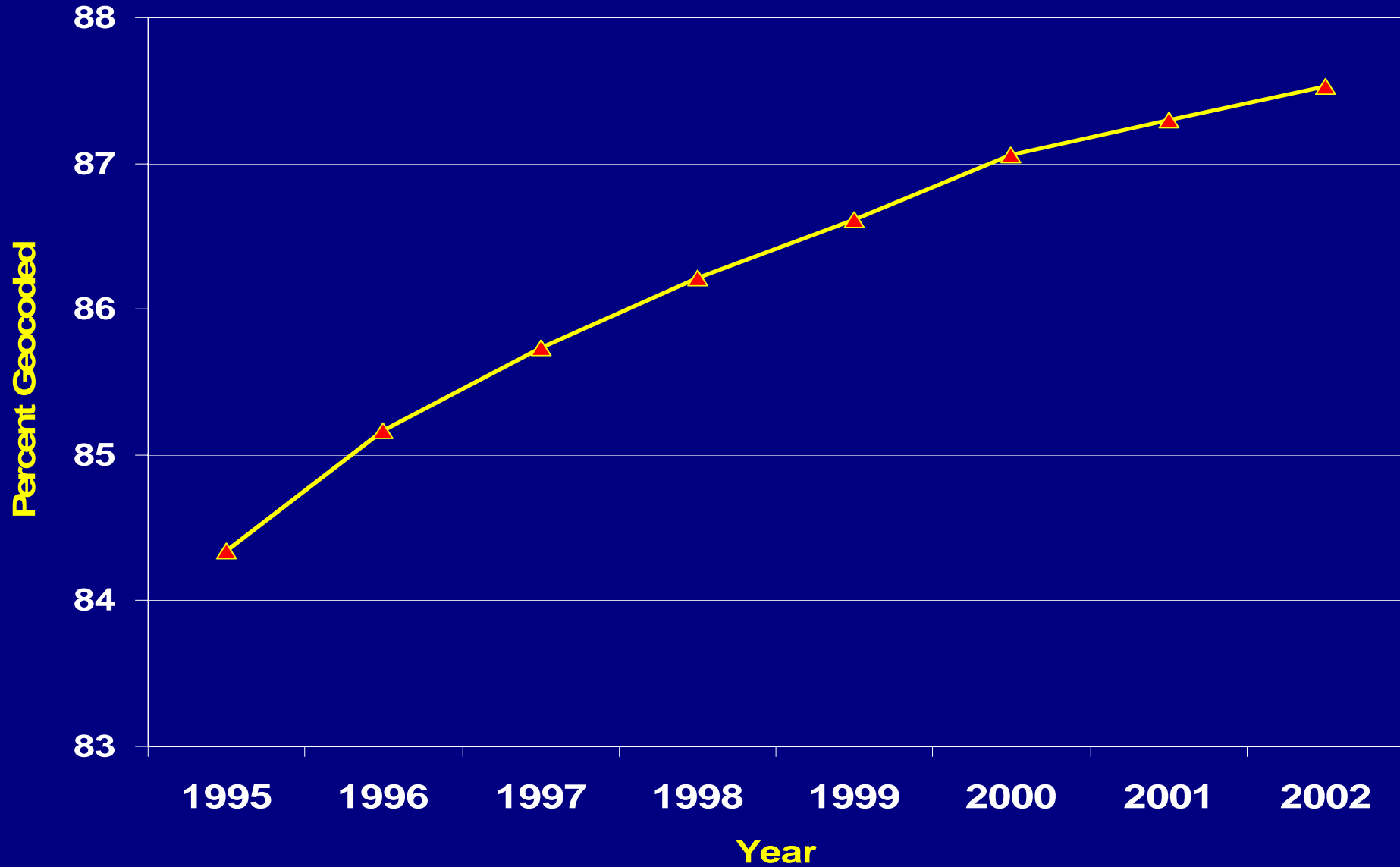


1990, 1991 ZIP Code Boundaries



Automated Geocoding Percent By Year

New York State Birth Records



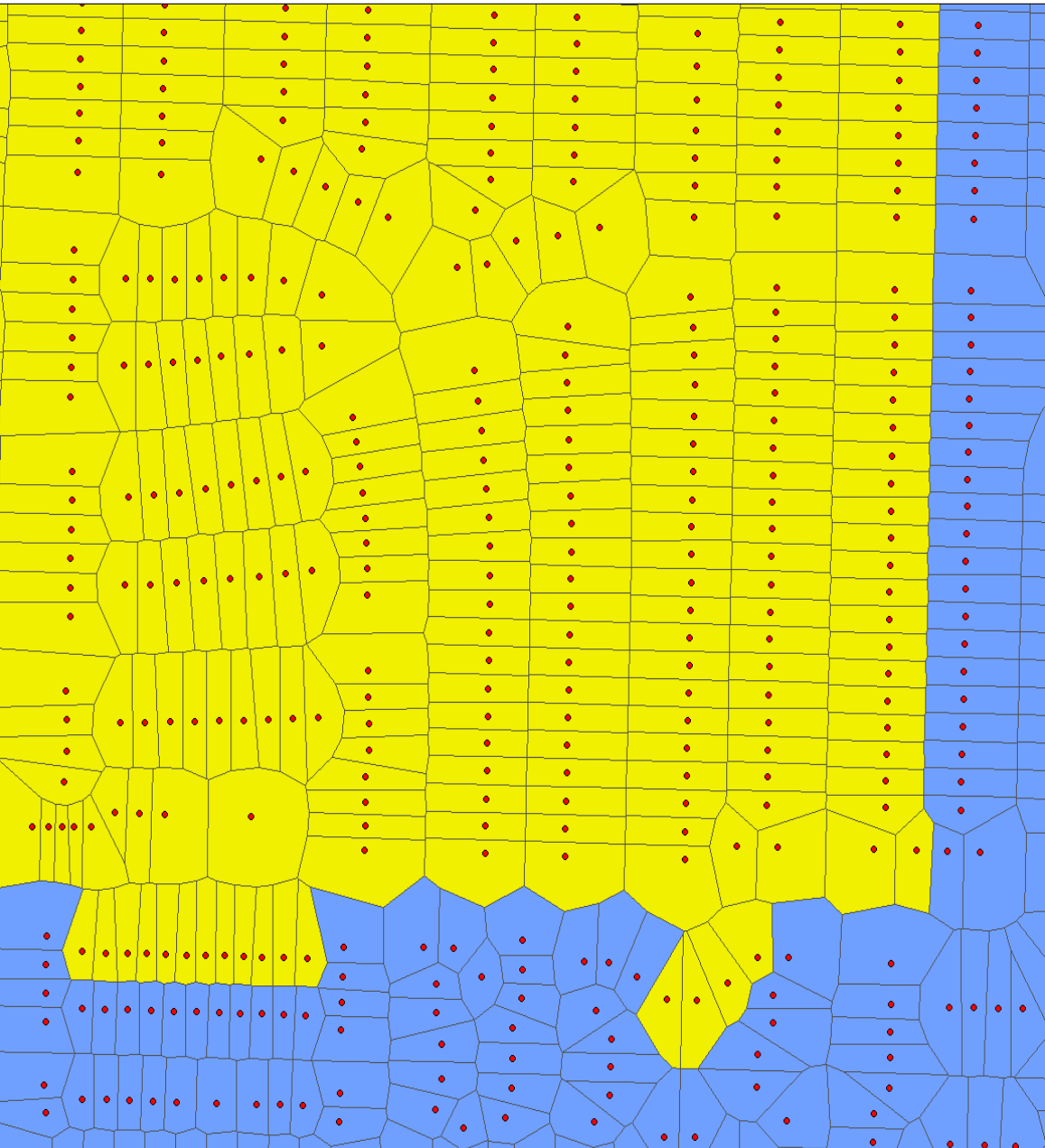
*Creation of
Zip Code
Boundaries Using
Real Property
Centroids*

LEGEND

● Real Property
Centroids

■ Voronoi Polygons
(14216)

■ Voronoi Polygons
(14214)



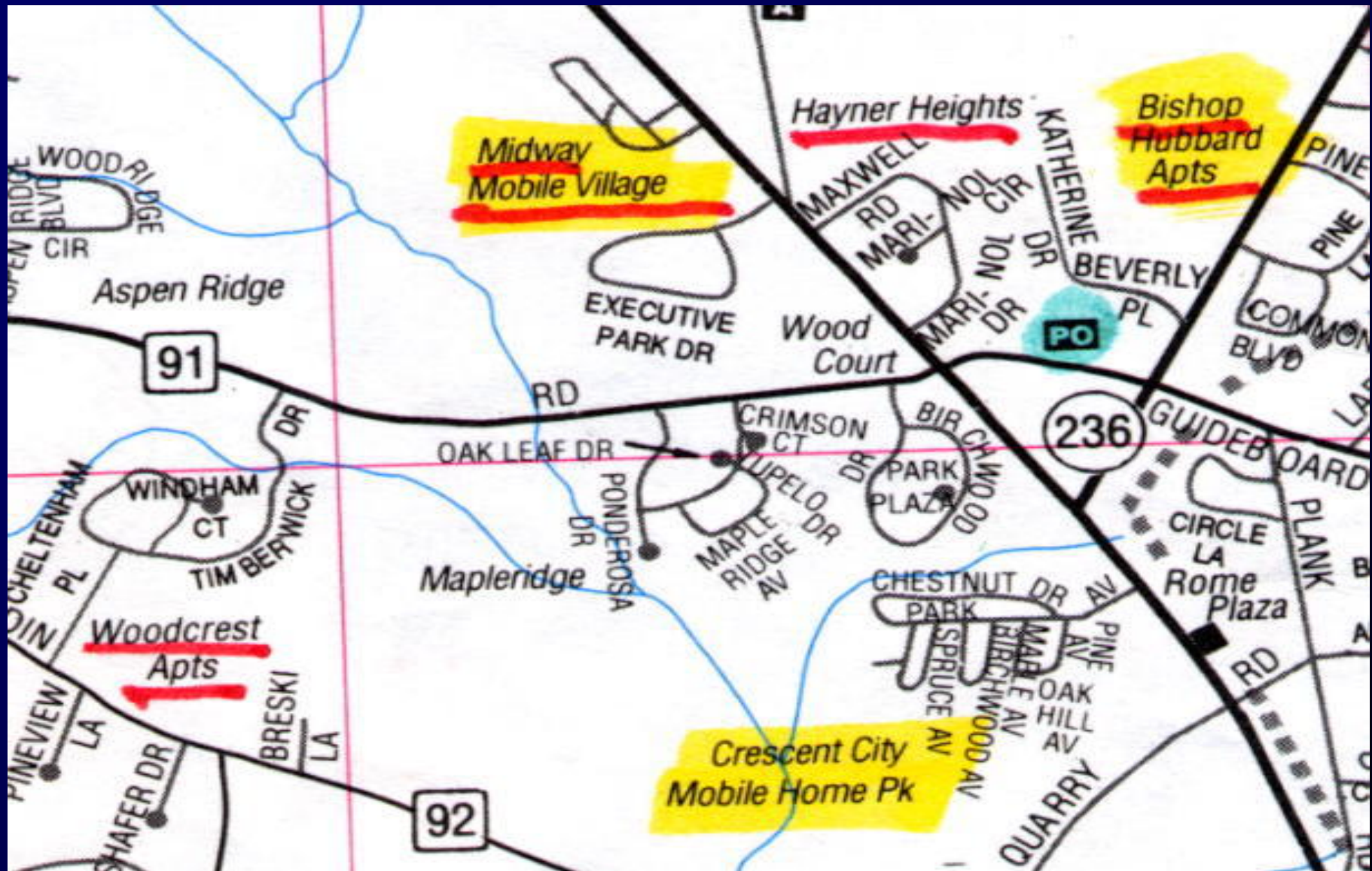
Other Sources

- Digital phone directories
- Property assessment data
- Aerial photography
- Paper Maps
- Internet

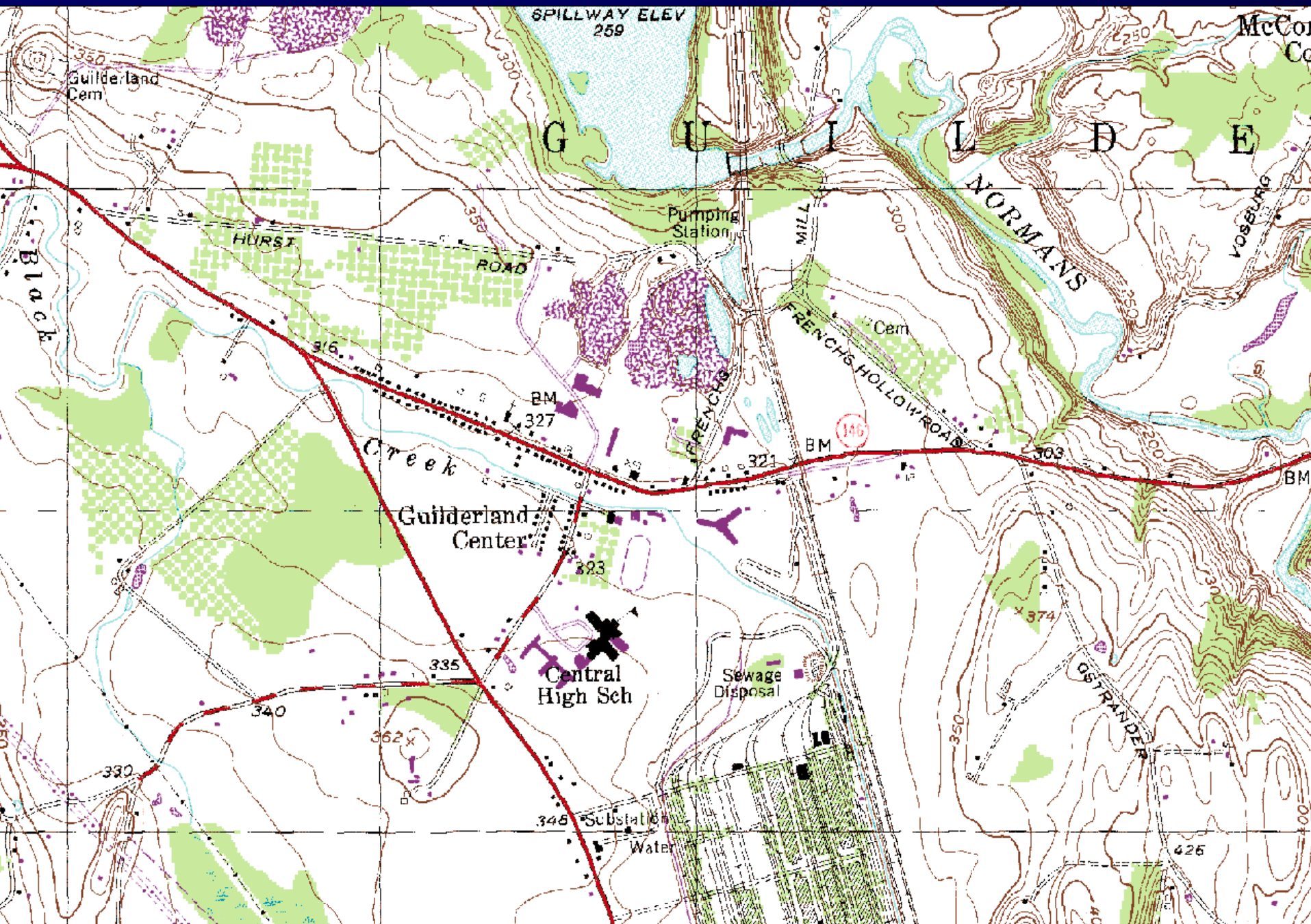
Other Sources

- Postal Service Files
- Department of Motor Vehicles
- City Directories
- E911

Finding mobile home parks & apartments



Finding things with USGS Digital TOPO Maps



Finding People through Dept. Motor Vehicles

QWS3270 EXTRA TN3270 Telnet - 150.142.96.60											
Host Edit Setup Help											
TODAY'S DATE: 1/31/2001 TIME: 13:51:83											
*RECORD EXPANSION FOR: TALBOT,THOMAS,0											
TALBOT,THOMAS,0						CLIENT ID#: 488842404					
118 LINCOLN AV						DOB: 03/27/1953 SEX: M					
ALTAMONT NY 12009						HEIGHT: 5-10 EYE COLOR: BLUE					
						COUNTY: ALBA					
						MI #: T01412 60043 702430-53					
LICENSE CLASS: *D*				STATUS: VALID				EXPIRATION: 03/27/2002			
***** ACTIVITY *****											
ACCIDENT PREVENTION COURSE COMPLETED ON: 02/24/2000											
UP TO 4 POINTS CAN BE REDUCED FROM TOTAL POINTS FROM 08/24/1998 - 02/24/2000											
N/A - NON 19-A DRIVER OR COURSE PRIOR TO 01/01/94											
*** ENTER NEXT FUNCTION CODE NEXT *** (RECORD CONTINUED ON FOLLOWING PAGE)											
NUM 13:48:29 IBM-3278-2											
ATTN	Clear	Erase EOF	Erase Input	Home	New Line	Tab	PA1	PA2	PA3	Reset	
PF1	PF2	PF3	PF4	PF5	PF6	PF7	PF8	PF9	PF10	PF11	PF12

Digital Phone Directories

MS-DOS Prompt - PD

Enter Name: TALBOT, THOMAS_

(press Tab to limit)=====UPSTATE NEW YORK AREA JAN - 1997=====

TALBOT, TAIN E 4635 N MAIN ST HEMLOCK, NY 14466	716-367-8845
TALBOT, TAMMY COPE CORNERS RD GILBERTSVILLE, NY 13776	607-783-2589
TALBOT, THEODORE STATE ROUTE 26 CARTHAGE, NY	315-493-3603
TALBOT, THOMAS 118 LINCOLN AVE ALTAMONT, NY 12009	518-861-5398
TALBOT, THOMAS 8397 COUNTY ROUTE 125 CHAUMONT, NY 13622	315-649-5246
TALBOT, THOMAS E MYERS MILLS RD NEW LISBON, NY 13415	607-263-5606
TALBOT, THOMAS L 3419 STONE QUARRY RD CAZENOVIA, NY 13035	315-655-2328
TALBOT, THOMAS N SR 5740 WOODSIDE RD SPRINGVILLE, NY 14141	716-592-7830
TALBOT, THOMAS Y 113 KATTELVILLE RD BINGHAMTON, NY 13901	607-648-9407
TALBOT, THOS F 8 CURRIE CT BALLSTON SPA, NY 12020	518-885-7134
TALBOT, THOS L 26 FORMAN ST CAZENOVIA, NY 13035	315-655-9216
TALBOT, THOS P 9 APPLEBLOSSOM LN VOORHEESVILLE, NY 12186	518-765-2509
TALBOT, TIMOTHY 233 W BANK ST ALBION,	
TALBOT, TIMOTHY 242 CAROLINA ST BUFFAL	
TALBOT, TODD COPES CORNERS RD GILBERTS	
TALBOT, V R 2 ALLMAN PL NEW HARTFORD,	
TALBOT, VALERIE 543 AVERILL AVE ROCHES	
TALBOT, VANE 39 SOUTH ST EDMESTON, NY	
TALBOT, VERA 2185 LANGFORD RD NORTH CO	
TALBOT, VINCENT A 159 LYNDALE AVE BUFF	
TALBOT, WALTER 135 BRUNSWICK RD TROY,	

TALBOT, THOMAS
118 LINCOLN AVE
ALTAMONT, NY 12009

518-861-5398 * Residence *

F1-Help Tab-Limit F4-?Index F5-Dial F6-Disc F8-Prt F9-Count Esc-Quit

Geocoding Results Differ between Commercial Products

- Black boxes with unknown algorithms
- Evaluate the software
 - match rates
 - and positional accuracy
- High match rates are not always better.
- “I can geocode everyone however they may not be in the right location”

The End